Supplement of: Organic aerosol source apportionment by offline-AMS over a full year in Marseille

C. Bozzetti et al.

Correspondence to: A. S. H. Prévôt (andre.prevot@psi.ch)

N. Marchand (nicolas.marchand@univ-amu.fr)

Table S1. Supporting measurements. Batch 1 denotes the set of filters collected during the yearly cycle from August 2011 to July 2012. Batch 2 indicated the set of filters collected during February 2011.

Analytical Method	Measured compounds / variable	Batch of filters
Tapered element oscillating microbalance equipped with a Filter Dynamic Measurement System	$PM_{2.5}$	1
IC (Jaffrezo et al., 1998)	SO_4^{2-}, NO_3^-, Cl^-, NH_4^+, Na^+, K^+, Ca^{2+}, Mg^{2+}, oxalate, malate, malonate and succinate	1,2
Thermal Optical Transmittance using Sunset Lab Analyzer (Birch and Cary, 1996)	CO_3^{2-}	1
Thermal Optical Transmittance using Sunset Lab Analyzer (EUSAAR2, Cavalli et al., 2010)	EC/OC	1,2
Water extraction Thermal Decomposition ND-IR determination using TOC analyzer (description in Bozzetti et al., 2016a)	WSOC	1

Water extraction Thermal Decomposition Chemilumenscence using TOC analyzer	Total nitrogen (TN)	1
(Chemical Derivatization) GC-MS (El Haddad et al., 2009; Favez et al., 2010)	**PAH**: phenantrene, anthracene, fluoranthene, acephenantrene, pyrene, benzo[a]anthracene, chrysene/Triphenylene, benzo[b,k]fluoranthene, benzo[j]fluoranthene, benzo-e-pyrene, benzo-a-pyrene, indeno[1,2,3 - cd]pyrene, dibenzo[a,h]anthracene, benzo - ghi - perylene	
(Chemical Derivatization) GC-MS (El Haddad et al., 2009; Favez et al., 2010)	**Alkanes**: octadecane (C18), nonadecane (C19), eicosane (C20), heneicosane (C21), docosane (C22), tricosane (C23), tetracosane (C24), pentacosane (C25), hexacosane (C26), heptacosane (C27), octacosane (C28), nonacosane (C29), triacontane (C30), untricontane (C31), dotriacontane (C32), tritriacontane (C33), tetratriacontane (C34), pentatriacontane (C35), hexatriacontane (C36)	1,2
(Chemical Derivatization) GC-MS (El Haddad et al., 2009; Favez et al., 2010)	**Hopanes**: 17α(H) - 21β(H)- norhopane (C29), 17α(H) - 21β(H)- hopane (C30), 17α(H) - 21β(H)-22R- homohopane (C31), 17α(H)-21β(H)-22S-homohopane (C31), 17α(H)-21β(H)-22S-bishomohopane (C32), 17α(H)-21β(H)-22R-bishomohopane (C32), 17α(H)-21β(H)-22S-trishomohopane (C33),	

	17α(H)-21β(H)-22R-trishomohopane (C33)			
(Chemical Derivatization) GC-MS (El Haddad et al., 2009; Favez et al., 2010)	**Cellulose and lignin pyrolysis products**: levoglucosan, vanilline, coniferaldehyde, syringaldehyde, acetosyringone, vanillic Acid, abietic Acid **Sterols**: cholesterol, stigmasterol, β - sitosterol			
(Chemical Derivatization) GC-MS (El Haddad et al., 2009; Favez et al., 2010)	**Fatty acids**: stearic acid, oleic acid, linoleic Acid, palmitic Acid **Phthalate esters**: di-ethyl phthalate, di-isobutyl phthalate, dibutyl phthalate, benzyl butyl phthalate, bis (2-ethyl hexyl) phthalate **Others**: pinonic acid			
ICP-MS (Chauvel et al., 2010; El Haddad et al., 2011).	Al, As, Ba, Ca, Cd, Ce, Co, Cr, Cs, Cu, Fe, K, La, Li, Mg, Mn, Mo, Na, Ni, Pb, Pd, Pt, Rb, Sb, Sc, Se, Sn, Sr, Ti, Tl, V, Zn, Zr			1,2
UPLC-ESI-ToF-MS (Iinuma et al., 2010).	4-methyl-5-nitrocatechol (major)/3-methyl-5-nitrocatechol (minor)	3-methyl-4-nitrocatechol	4-nitrocatechol	1

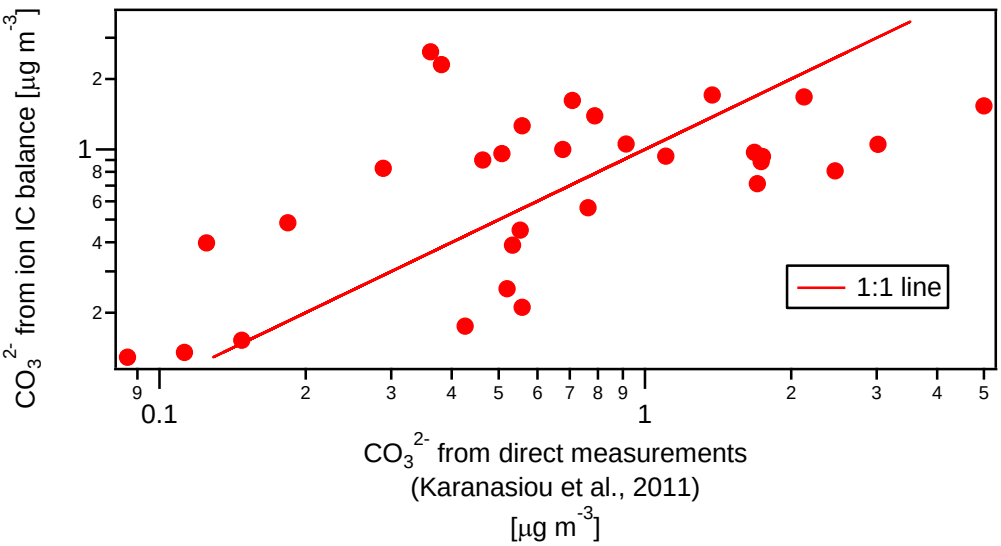


Figure S1. Scatter plot of CO_3^{2-} measurements (Karanasiou et al., 2011) *vs.* CO_3^{2-} estimates from the IC ion balance. The CO_3^{2-} molar concentration from ion balance was estimated as the difference between the equivalents of cations (Ca^{2+}, K^+, NH_4^+, Na^+, Mg^{2+}) and anions (NO_3^-, SO_4^{2-}, Cl^-).

HCO_3^- correction of offline-AMS spectra.

As mentioned in the manuscript, the measured pH of the filter extract never exceeded 8, indicating the absence of CO_3^{2-} in solution, and that we can assume that water-extracted CO_3^{2-} is present as HCO_3^-. Considering all the measured HCO_3^- as deriving from $Ca(HCO_3)_2$ or from $NaHCO_3$, none of the liquid extracts exceeded the $Ca(HCO_3)_2$ or $NaHCO_3$ saturation concentrations at 20°C. Even considering all the measured CO_3^{2-} to be in the $CaCO_3$ form (however we can exclude this assumption as the pH of the liquid extracts was always < 8), only one filter extract showed a CO_3^{2-} concentration exceeding the $CaCO_3$ saturation concentration. Therefore we can assume all the estimated CO_3^{2-} (from IC ion balance) to be solubilized and in the HCO_3^- form. This would be our best estimate of the HCO_3^- water-soluble concentration. In the following we also assess the sensitivity of the source apportionment results on the HCO_3^- correction of the PMF input matrices (described hereafter) by performing a source apportionment without HCO_3^- corrections.

The HCO_3^- correction was implemented by estimating the HCO_3^- relative ionization efficiency (RIE) with respect to NO_3^-. We measured nebulized and size-selected $NaHCO_3$ particles (400 nm mobility diameter, using a differential mobility analyzer, DMA) in the

AMS. From the particle-ToF signal (pToF, Fig. S2) of the AMS we determined a $NaHCO_3$ Jayne shape factor *S* of 0.9±0.1 (Jayne et al. 2000, variability from multiple $NaHCO_3$ injections), defined as:

$$S = \frac{d_{va}}{d_m} \cdot \frac{\rho_0}{\rho_m} \quad \text{(S1)}$$

Here d_{va} denotes the aerodynamic diameter under vacuum measured by the AMS and d_m the DMA mobility diameter (DeCarlo et al., 2004). ρ_0 is the standard density of 1 g cm^{-3} and ρ_m represents the $NaHCO_3$ density (2.2 g cm^{-3}). Perfectly spherical particles are characterized by an *S* value of 1, in our case we observed *S* values not significantly different from 1 (within our uncertainties) for standard $NaHCO_3$ injections, indicating that there is no reason to consider non-spherical particles. Note that previous DMA and pToF calibrations were conducted using polystyrene (PSL) spherical particles with known diameters.

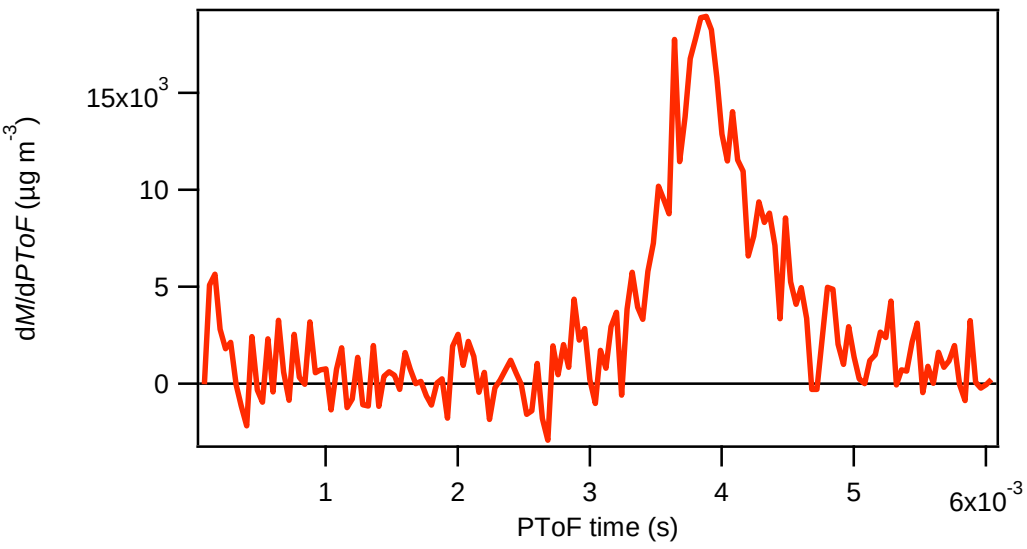


Figure S2. $NaHCO_3$ pToF signal.

Data from $NaHCO_3$ nebulization were collected using the single particle (brute force single particle (BFSP), Drewnik et al., 2004) AMS operating mode, tracing the HCO_3^- signal at *m/z* 44. By using the $NaHCO_3$ particle density, and the newly determined Jayne shape factor (0.9±0.1) we calculated the number of HCO_3^- ions per particle. The HCO_3^- ionization efficiency ($IE_{HCO_3^-}$) was calculated by dividing the number of CO_2^+ ions detected per particle by the number of $NaHCO_3$ molecules per particle ($IPP_{HCO_3^-}$)) and by the *m/z* 44 fractional contribution (*f*44) to the HCO_3^- spectrum (*f*44 = 0.44) in order to account for the contribution

of other fragments to the HCO3- spectrum. The HCO_3^- relative IE ($RIE_{HCO_3^-}$) in comparison to NO_3^- was determined as follows:

$$RIE_{HCO_3^-} = \frac{IE_{\frac{m}{z}44,HCO_3^-}}{IE_{NO_3^-}} \cdot \frac{ionization\ cross\ section\ (NO_3^-)}{ionization\ cross\ section\ (HCO_3^-)} \quad \text{(S2)}$$

Here the molecular weights of NO_3^- and HCO_3^- were used as proxies for the corresponding ionization cross sections. $RIE_{HCO_3^-}$ was determined to be 1.4±0.2 (0.2 is the variability from multiple HCO_3^- nebulizations (n=3) which includes the ion per particle counting uncertainty), which is not statistically different from the standard *RIE* assumed for organics (RIE_{org} = 1.4).

Water-soluble mass spectra were corrected as described hereafter. Inputs for this correction are:

- measured water-soluble normalized AMS spectra $(\overrightarrow{AMS})_{norm,i}$ and corresponding OM:OC ratios $\left(\frac{OM}{OC}\right)'_i$,
- $WSOC_i$ measurements (TOC analyzer),
- $HCO_3^-{}_i$ estimates from IC ion balance,
- HCO_3^- normalized AMS spectrum $\overrightarrow{HCO_3^-}_{norm}$ as measured from $NaHCO_3$ solution nebulization, and corresponding $\left(\frac{HCO_3^-}{C_{HCO_3^-}}\right)$ ratio determined from the HCO_3^- AMS spectrum, (4.01).

No correction for gaseous CO_2 was applied to the $\overrightarrow{HCO_3^-}_{norm}$ spectrum as the CO_2 and HCO_3^- fragmentation is supposed to be the same due to the HCO_3^- thermal decomposition into CO_2 and H_2O onto the vaporizer. Moreover, the fragments deriving from the water fragmentation (O^+, OH^+, and H_2O^+) do not introduce differences into the CO_2 and HCO_3^- spectra because their intensities were estimated from the CO_2^+ fragment according to the standard AMS fragmentation table (Aiken et al., 2008).

For a generic filter sample *i*, the measured $(OM/OC)_i$' ratio represents a linear combination of $(OM/OC)_i$ contributions deriving from $HCO_3^-{}_i$ and from $WSOM_i$. Considering the statistically not different *RIE*s of organics and HCO_3^-, we can assume the organics and HCO_3^- AMS response to be not different. In the same way, considering internally mixed particles from filter extracts nebulization, we assumed equal CE for both WSOM and HCO_3^-.

$$\left(\frac{OM}{OC}\right)'_i = \left(\frac{WSOM+HCO_3^-}{WSOC+C_{HCO_3^-}}\right)_i \quad \text{(S3)}$$

where $C_{HCO_3^-}$ represents the C concentration deriving from HCO_3^- as measured by the AMS, calculated from the $HCO_3^-{}_i$ absolute concentrations (from IC ion balance) divided by the $\left(\frac{HCO_3^-}{C_{HCO_3^-}}\right)$ ratio determined from the AMS HCO_3^- spectrum. Similarly to $\left(\frac{OM}{OC}\right)'_i$, also the AMS mass spectral fingerprint can be considered as the sum of $WSOM_i$ and $HCO_3^-{}_i$, therefore the normalized blank-subtracted AMS spectra $(\overrightarrow{AMS})_{norm,i}$ and corresponding errors $\sigma(\overrightarrow{AMS})_{norm,i}$ were rescaled to the sum of $WSOM_i$ and $HCO_3^-{}_i$ calculated as the sum of $WSOC_i$ (from TOC analyzer) and $C_{HCO_3^-,i}$ (ion balance) multiplied by $\left(\frac{OM}{OC}\right)'_i$

$$(\overrightarrow{AMS})_i = (\mathrm{WSOC}_i + C_{HCO_3^-,i}) \cdot \left(\frac{OM}{OC}\right)'_i \cdot (\overrightarrow{AMS})_{norm,i} \quad \text{(S4)}$$

By dividing numerator and denominator of $\frac{WSOM}{WSOC+C_{HCO_3^-}}$ by WSOC and by dividing numerator and denominator of $\frac{HCO_3^-}{WSOC+C_{HCO_3^-}}$ by $C_{HCO_3^-}$, we can express Eq. S3 as:

$$\left(\frac{OM}{OC}\right)'_i = \frac{(\frac{OM}{OC})_{WSOM,i}}{1+\frac{C_{HCO_3^-}}{WSOC}} + \frac{\left(\frac{HCO_3^-}{C_{HCO_3^-}}\right)}{1+\frac{WSOC}{C_{HCO_3^-}}} \quad \text{(S5)}$$

From Eq. S5 we can derive $(\frac{OM}{OC})_{WSOM,i}$. The time dependent WSOM concentration was therefore calculated as $(\frac{OM}{OC})_{WSOM}$·WSOC.

For a generic filter sample *i*, the HCO_3^- AMS signature $(\overrightarrow{HCO_3^-})_i$ can be determined as:

$$(\overrightarrow{HCO_3^-})_i = C_{HCO_3^-}{}_i \cdot \left(\frac{HCO_3^-}{C_{HCO_3^-}}\right) \cdot \overrightarrow{HCO_3^-}_{norm} \quad \text{(S6)}$$

where $\overrightarrow{HCO_3^-}_{norm}$ represents the normalized HCO_3^- AMS spectrum derived from standard injection. To derive the AMS signal purely generated by WSOM $(\overrightarrow{WSOM})_i$, we subtracted $(\overrightarrow{HCO_3^-})_i$ (calculated as in Eq. S6) from $(\overrightarrow{AMS})_i$ (calculated as in Eq. S4).

$$(\overrightarrow{WSOM})_i = (\overrightarrow{AMS})_i - (\overrightarrow{HCO_3^-})_i \quad \text{(S7)}$$

The CO_3^- concentration uncertainty ($\sigma_{CO_3^-}$) was estimated by propagating the error for all the ions (measured by IC) used to estimate the CO_3^- concentration from the ion balance (Ca^{2+}, Mg^{2+}, K^+, Na^+, NH_4^+, Cl^-, NO_3^-, SO_4^{2-}). For a generic ion *s*, the errors ($\sigma_{i,s}$) were estimated by propagating the detection limits $(DL)_s$ and the relative repeatability $(RR)_s$ multiplied by the ion concentration according to Eq. (S8) (Rocke and Lorenzato, 1995):

$$\sigma_{i,s}=\sqrt{DL_s{}^2+(x_i\cdot RR)_s{}^2} \quad \text{(S8)}$$

We assumed for each ion an uncertainty deriving from IC detection limit. On average this CO_3^{2-} uncertainty was 28%. The uncertainty associated to $(\overrightarrow{HCO_3^-})_i$ was instead estimated as:

$$\sigma(\overrightarrow{HCO_3^-})_i = \sigma_{CO_3^-}\cdot\frac{AW_C}{FW_{CO_3^{2-}}}\cdot\left(\frac{HCO_3^-}{C_{HCO_3^-}}\right)\cdot\overrightarrow{HCO_3^-}_{norm}. \quad \text{(S9)}$$

where AW_C is the carbon atomic weight and $FW_{HCO_3^-}$ represents the CO_3^{2-} molecular weight. The final $WSOM_i$ mass spectral uncertainty $\sigma(\overrightarrow{WSOM})_i$ error was estimated by summing under quadrature ($\sigma(\overrightarrow{HCO_3^-})_i$) and the error associated to the total AMS signal rescaled for the sum of WSOM and HCO_3^-.

$$\sigma\,(\overrightarrow{WSOM})_i = \sqrt{\left(\sigma\left(\overrightarrow{HCO_3^-}\right)_i\right)^2+\left(\sigma\left(\overrightarrow{AMS}\right)_i\right)^2} \quad \text{(S10)}$$

Influence of the HCO_3^- correction on the offline-AMS source apportionment results.

As we have mentioned in the main text, Section 2.4, the offline AMS measurements of the WSOC are influenced by the presence of inorganic carbonates. This influence has been corrected using the carbonate mass estimated from the ion balance obtained by IC measurements. If this correction is not applied, PMF separates an additional factor with a highly oxidized fingerprint similar to inorganic carbonate and whose time series strongly correlates with that of Ca^{2+}. In the following we compare the carbonate mass estimated from IC measurements and from PMF, and assess the influence of the correction applied on the estimation of the different factors.

We performed a source apportionment on the non HCO_3^- corrected input matrices. Input matrices were scaled to $WSOM_i$ concentration calculated as $WSOC_i$ multiplied by $\left(\frac{OM}{OC}\right)'_i$. We explored a 6-factor solution where the additional separated factor was attributed to inorganic dust. For a generic water soluble K factor (WSKOA) and a generic time element *i*, the corresponding water soluble OC concentrations $(WSKOC)_i$ were multiplied by a factor $(1/(1-f_{\text{inorganic dust}}))_i$, where $f_{\text{inorganic dust}}$ represents the relative contribution of the water soluble inorganic dust factor as obtained from PMF. The application of the $(1/(1-f_{\text{inorganic dust}}))_i$ factor enables rescaling the sum of the water soluble OC concentrations of the five organic factors (WSHOC, WSBBOC, WSINDOC, WSCOC, WSOOC) to the measured WSOC

concentration, under the assumption of no organic contributions to the inorganic dust factor. The accuracy of this assumption is discussed in the following.

In total 240 PMF runs were performed. PMF solutions were retained according to the acceptance criteria 1-6 listed in Section 2.4. We found that primary sources (WSHOC, WSCOC, WSBBOC, and WSINDOC) showed statistically not different contributions with the offline-AMS source apportionment conducted on HCO_3^--corrected spectra. However, for some of the retained PMF solutions, the inorganic dust-related factor tended to strongly mix with WSOOC because the two factors were characterized by the highest fCO_2^+ (in this case none of the factors showed a strong correlation with HCO_3^- concentrations). In order to retain the solutions that best resolved the inorganic dust from the WSOOC factor we introduced two further acceptance criteria:

1) Significantly positive *R* between inorganic dust and HCO_3^-.
2) Inorganic dust correlation with HCO_3^- (*R*) significantly higher than the correlation between WSOOC and HCO_3^-.

Consequently, half of the solutions retained according to criteria 1-6 were discarded.

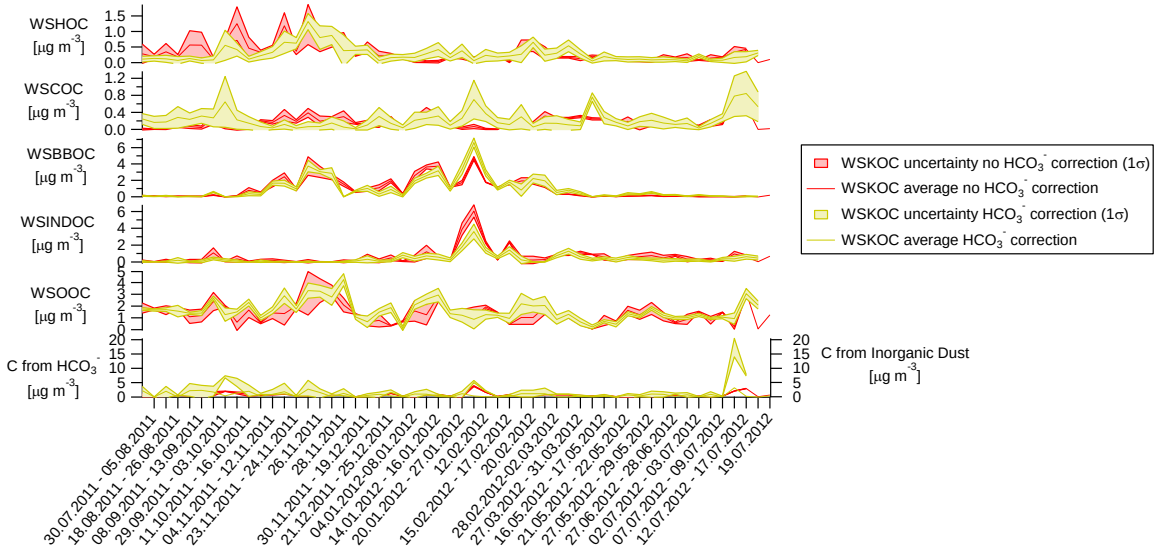


Figure S3. Source apportionment results obtained with and without HCO_3^- correction.

WSHOC, WSBBOC, WSOOC, and WSINDOC showed not statistically different concentrations (99% confidence interval) with and without the HCO_3^- correction. The WSCOC factor instead revealed statistically different concentrations within a 99% confidence interval, but only for 12% of the data points. Overall the PMF estimate of the C from

inorganic dust was higher than the C estimate from HCO_3^- derived from the IC ion balance, and was more uncertain. This can be explained by an imperfect separation from other factors (especially WSOOC). The difficult separation between WSOOC and inorganic dust hampered an accurate post-PMF HCO_3^- correction. Therefore we opted to show in the manuscript source apportionment results obtained performing a pre-PMF HCO_3^- correction of the OA input matrices, but we note that this correction while uncertain does not have a significant effect on factor retrieval.

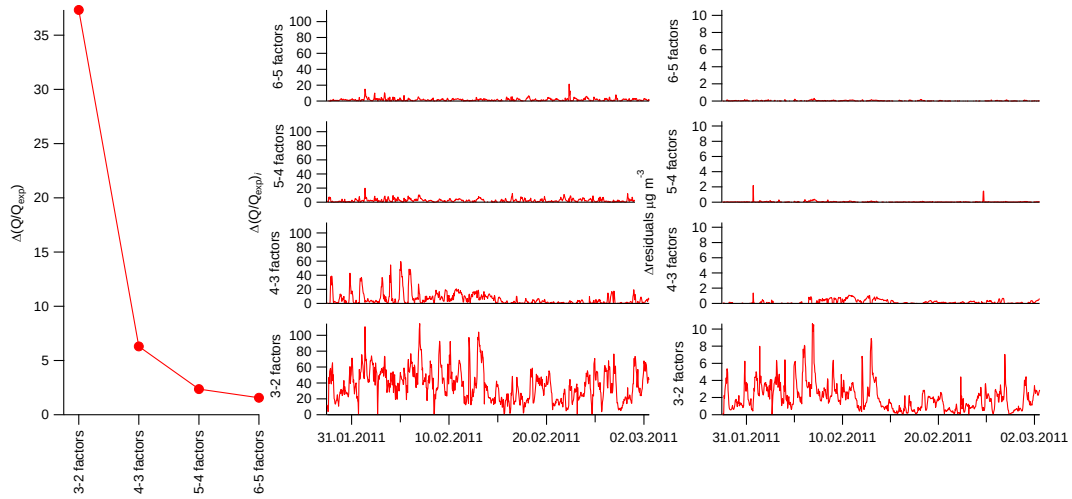


Figure S4. Online-AMS PMF $\Delta Q/Q_{exp}$ analysis. In this study, a large Q/Q_{exp} decrease could be observed up to 4 factors. The 4-factor solution enabled resolving BBOA which is mixed with OOA in the 3-factor solution. Increasing further the number of factors provided only small additional contributions to the explained variability, resulting in a splitting of HOA in the 5-factor solution, and BBOA in the 6-factor solution; however the newly separated factors could not be attributed to specific different sources. In terms of residuals, a clear removal of the structure can be observed up to 4 factors.

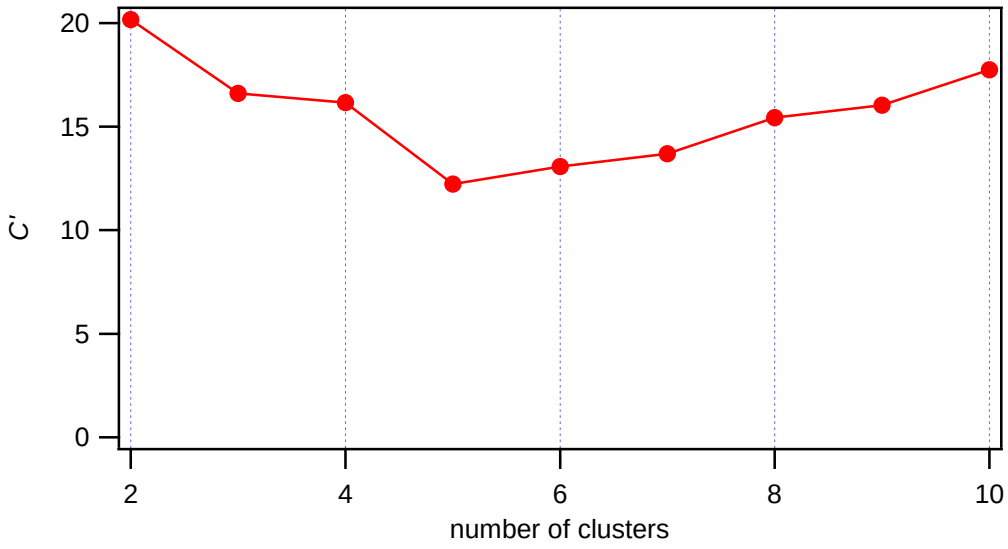


Figure S5. Number of cluster selection: *C* values (Eq. 7) as a function of the number of clusters.

Best cluster selection

From the a-value sensitivity analysis, 121 solutions were obtained, the diurnal time series of which were clustered using a k mean clustering approach. The clusters were then filtered based on the cosine similarity of the HOA, COA, and BBOA average cluster mass spectra with the average mass spectra reported in the AMS literature for the same factors (Crippa et al., 2013b, Mohr et al., 2012 and 2009, Bruns et al., 2015, Docherty et al., 2011, Setyan et al., 2012, He et al., 2010). Given two vectors (in our case mass spectra) *A* and *B* with *n* elements each, the cosine similarity is defined as:

$$cosine\ similarity = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \cdot \sqrt{\sum_{i=1}^{n} (B_i)^2}} \quad \text{(S11)}$$

Eq. S11 returns values between -1 and 1, with 1 meaning collinearity, and 0 orthogonality. Because of the different HR fittings performed in different works, we considered only fragments in common with our HR fit for comparison. Within the common variables, we selected a subset of fragments characterized by small variability in comparison to the average value ($S/N>2$) for at least one average mass spectrum among COA, HOA, and BBOA. Here *S* denotes the average literature value for a certain fragment, and N represents its standard deviation. This selection was performed in order to choose the most stable and therefore certain fragments characterizing the reference spectra. Following this step, 95, 92, and 87% of the HOA, COA, and BBOA mass was retained for the average reference mass spectra,

respectively. For our dataset, depending on the cluster we selected 91-93% of the COA mass, 84-86% for HOA, and 91-93% for BBOA. We explored the deviation of the excluded fragments from the literature values by checking whether their relative contributions to the factor mass spectra were within the literature range (maximum and minimum). For this comparison we calculated the average HOA, COA, BBOA, and OOA spectra for each of the 5 clusters. This average includes all the PMF factor profiles attributed to a specific cluster for the 100 random initiations of the *k*-mean algorithm. Overall, depending on the cluster, only 1-2% of the total HOA mass wasn't included within the literature range, 2-5% for COA, while for BBOA the value was smaller than 0.5%. These diagnostics highlight the relevance in terms of mass of the fragments retained for the cosine similarity comparison.

For each source (BBOA, COA, HOA), individual literature spectra were compared with the corresponding average literature spectrum in order to estimate the minimum cosine similarity value characterizing the average profiles. This minimum value is used as a threshold above which a spectrum can be considered not different from literature profiles within 1σ. The obtained cosine similarities were 0.965±0.008 for HOA, 0.96±0.05 for COA, and 0.94±0.06 for BBOA. Therefore the minimum cosine similarities to define a mass spectrum as not statistically different from the average reference spectra were 0.957 for HOA, 0.91 for COA, and 0.88 for BBOA.

We note that that reference HOA and COA spectra are less variable (0.965±0.008 and 0.96±0.05 respectively) than BBOA (0.94±0.06). This is probably due to the different fuels and burning conditions characterizing the different BBOA ambient and chamber profiles.

Subsequently we checked whether the average HOA, COA, and BBOA reference profiles showed statistically different spectra with each other. This was tested by calculating the cosine similarity between the average HOA, COA, and BBOA literature profiles and all the aforementioned profiles reported in literature (Crippa et al., 2013b, Mohr et al., 2012 and 2009, Bruns et al., 2015, Docherty et al., 2011, Setyan et al., 2012, He et al., 2010) for factors of a different type (e.g. average HOA *vs.* all COA). Our results indicate that HOA, BBOA, and COA average literature mass spectra show statistically different profiles within 1σ – average cosine similarities: HOA - BBOA: 0.5±0.1; HOA - COA: 0.83±0.05, COA - BBOA: 0.6±0.1; COA – HOA: 0.83±0.08, BBOA - COA: 0.6±0.1; BBOA - HOA: 0.5±0.2. This means that PMF factors can be identified based on the analysis approach we have adopted, as factor mass spectra are characteristic: i.e. differences between mass spectra pertaining to the

same factor are significantly smaller then differences between mass spectra related to different factors.

In order to select the best clusters we determined the cosine similarity of the average cluster mass spectra with the average reference profiles. A cluster was retained if the HOA, COA, and BBOA average cluster spectra were not statistically different from the corresponding average reference profiles. However the average cluster mass spectra are also characterized by an uncertainty deriving from the *k*-mean algorithm random initiation. To calculate this uncertainty, we generated 100 random cluster profiles by randomly varying the average cluster mass spectra within the corresponding standard deviation (calculated as the standard deviation of the cluster profiles obtained initiating the *k*-means algorithm 100 times) assuming a normal distribution of the error. Each randomly generated profile was compared with the average reference spectrum by calculating the corresponding cosine similarity. This provides the cosine similarity uncertainty of an average cluster spectrum with the literature average reference profile. From the comparison of average cluster mass spectra with reference spectra, we observed that HOA and BBOA showed statistically not different fingerprints with the corresponding average reference profiles (within 1σ) for all clusters, while COA showed a statistically different mass spectrum with respect to the average profile for cluster 5, which therefore was not retained for further analysis. For cluster 4, although COA average spectrum was statistically not different from the average COA reference profile within 1σ, the mass spectrum was not statistically different either from the HOA average mass spectrum within 1σ, suggesting a certain mixing of the two sources, therefore also cluster 4 was rejected (Table S2).

Table S2. Cosine similarity between COA, HOA, and BBOA average cluster spectra with the corresponding reference profiles from literature (average of the profiles reported in: Crippa et al., 2013b, Mohr et al., 2012 and 2009, Bruns et al., 2015, Docherty et al., 2011, Setyan et al., 2012, He et al., 2010). Threshold cosine similarity indicates the minimum cosine similarity value which defines a cluster profile as not statistically different from the reference profiles. Highlighted values indicate cluster profiles not statistically different from the reference profiles.

	threshold	0.957	0.91	0.89
		HOA reference spectrum	COA reference spectrum	BBOA reference spectrum
HOA	cluster1	0.99188 (3)	0.6931 (1)	0.4677 (2)

	cluster2	0.99532 (2)	0.69320 (8)	0.5237 (1)
	cluster3	0.99177 (6)	0.6752 (2)	0.5254 (3)
	cluster4	0.99025 (9)	0.6753 (4)	0.5448 (5)
	cluster5	0.99674 (4)	0.7079 (2)	0.5195 (4)
COA	cluster1	0.8373 (1)	0.99208 (5)	0.6453 (4)
	cluster2	0.85270 (9)	0.98997 (2)	0.62187 (8)
	cluster3	0.8861 (2)	0.98598 (4)	0.6045 (2)
	cluster4	0.9565 (8)	0.9648 (3)	0.6219 (8)
	cluster5	0.80 (1)	0.89 (1)	0.855 (2)
BBOA	cluster1	0.594 (1)	0.588 (1)	0.9687 (5)
	cluster2	0.7134 (3)	0.6874 (1)	0.94746 (7)
	cluster3	0.6945 (2)	0.6739 (2)	0.95441 (6)
	cluster4	0.578 (1)	0.598 (1)	0.96752 (5)
	cluster5	0.713 (2)	0.682 (2)	0.943 (1)

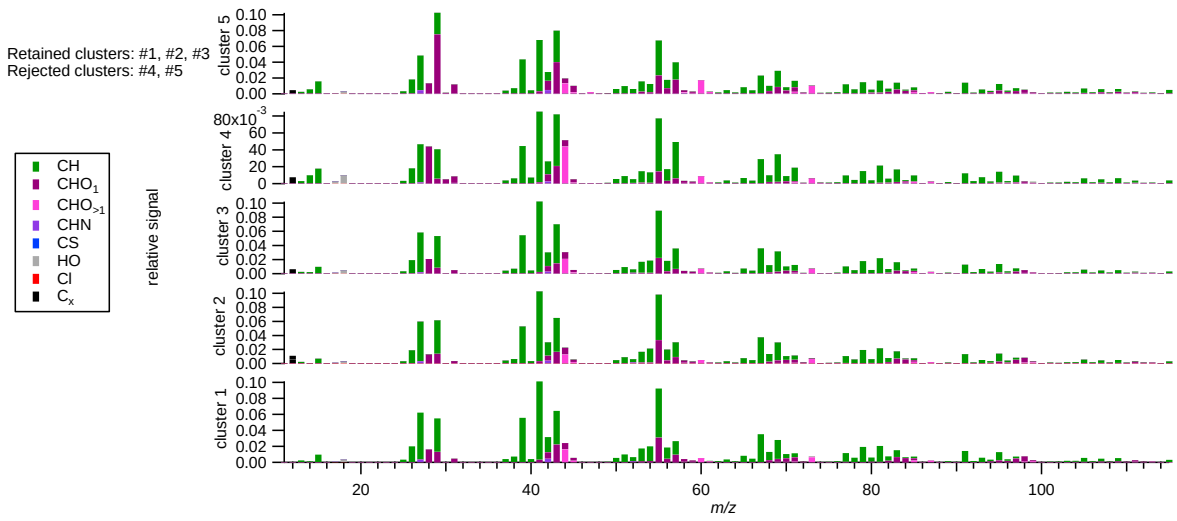


Figure S6. Average COA cluster spectra.

These results are also reflected by the high average COA diurnal pattern correlation with NO_x (typical traffic tracer) for cluster 4 (R=0.64), while for cluster 5 the average COA diurnal pattern correlates well with AMS-PAH (highly correlated with BBOA) and fragment $fC_2H_4O_2^+$ diurnals (R = 0.94 and 0.98 respectively), suggesting a certain mixing with BBOA (Fig. S7). For clusters 1-3, the COA diurnal correlation with NO_x and AMS-PAH was smaller than for clusters 4-5, and smaller than the correlation of HOA with NO_x and BBOA with AMS-PAH indicating a good COA separation from HOA and NO_x (Table S2).

Table S3. Correlation of COA cluster diurnals with NO_x, AMS-PAH, *m/z* 60, and suboptimal clusters. High correlations with NO_x suggest possible mixing between COA and HOA; high correlations with AMS-PAH and *m/z* 60 are suggestive of possible mixings between BBOA and COA; high correlations with clusters 4 and 5 are probably indicative of uncertain attribution of the PMF runs between the optimal and suboptimal clusters.

COA diurnals					
Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Pearson correlation coefficient
0.84	0.52	0.48	0.72	0.94	AMS-PAH
0.19	0.34	0.57	0.64	0.09	NO_x
0.82	0.42	0.30	0.54	0.98	*m/z* 60
0.77	0.83	0.93	1	0.61	cluster 4 COA diurnal
0.91	0.56	0.42	0.61	1	cluster 5 COA diurnal

Because of the small variability and the relatively high correlation coefficients between HOA and NO_x, and between BBOA and AMS-PAHs among solutions belonging to the retained clusters (*R* ranging between 0.76-0.79 and 0.92-0.93 respectively, Fig. S9), we did not select additional acceptance criteria based on HOA and BBOA temporal trends.

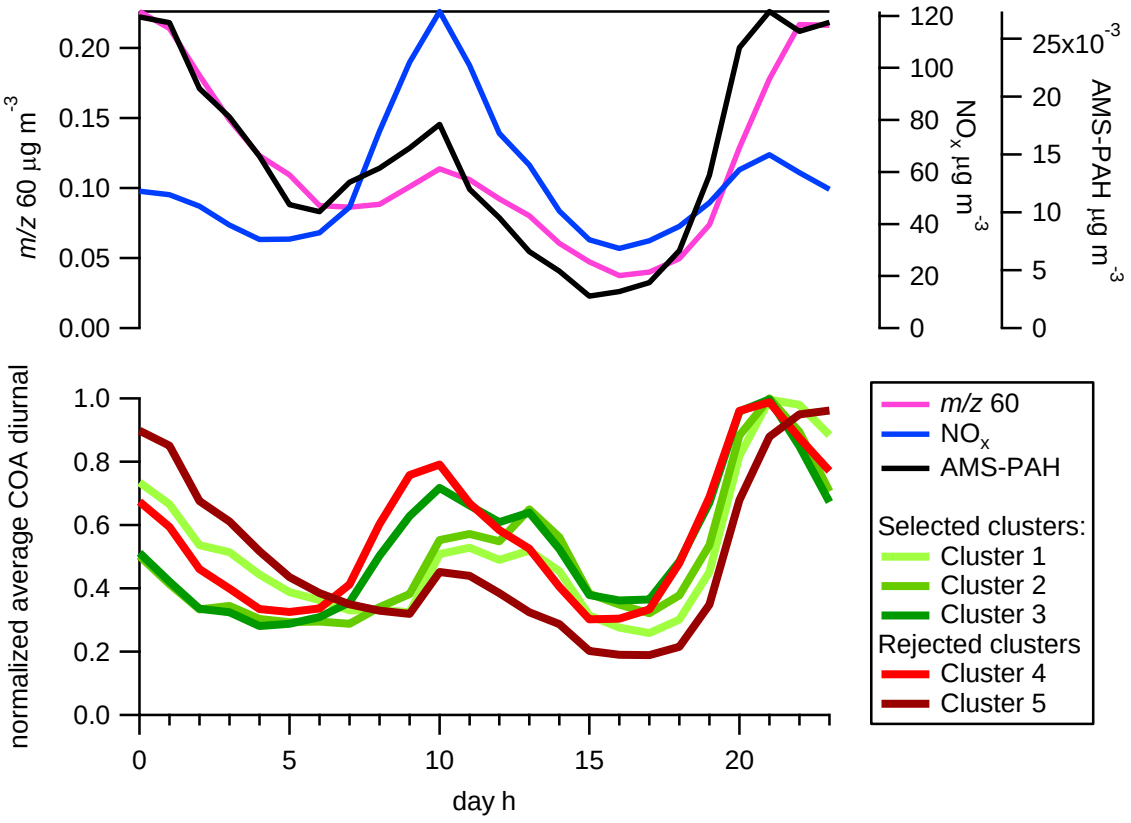


Figure S7. Average COA diurnal cycles for the different clusters.

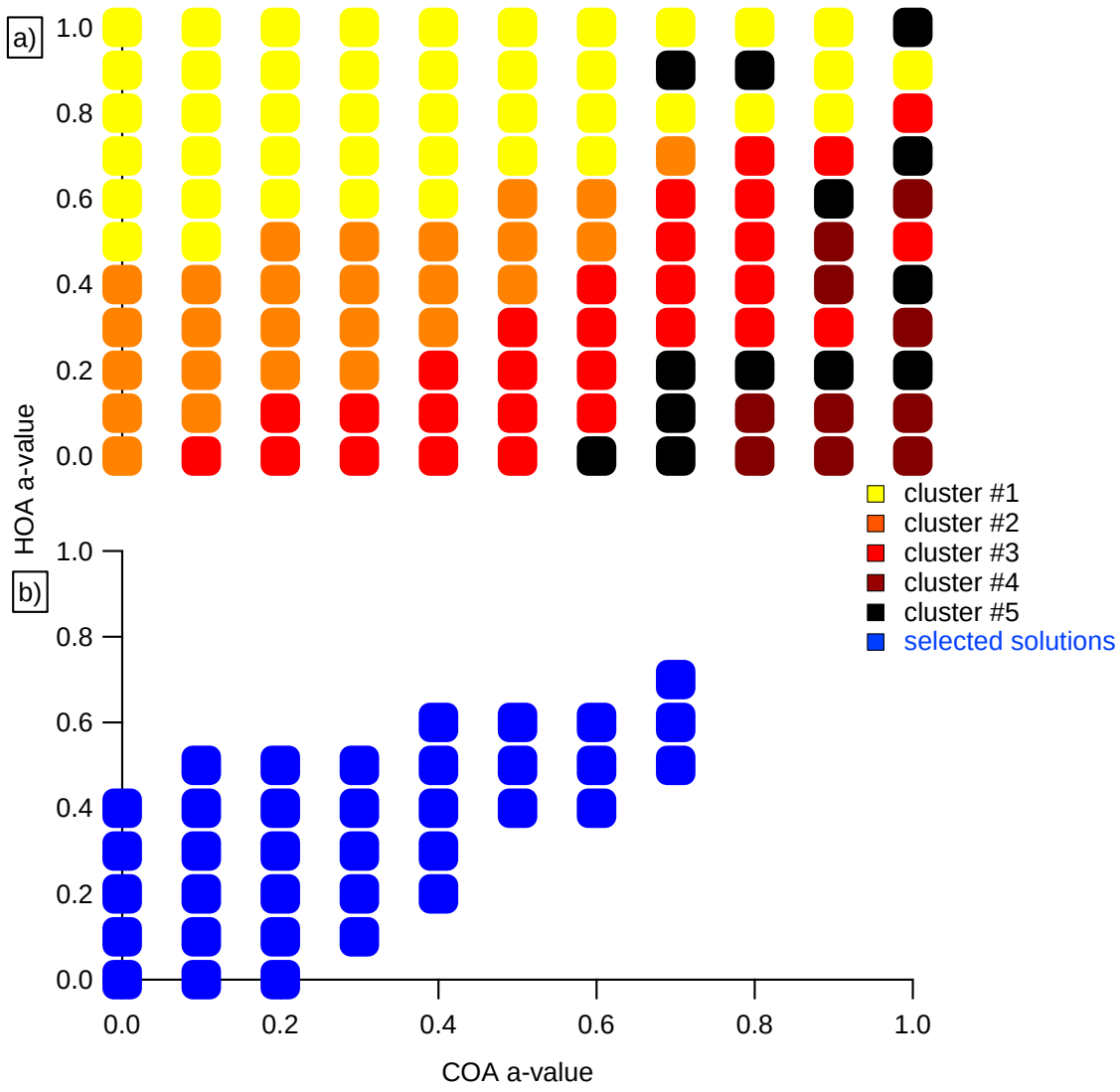


Figure S8. *a*-value sensitivity analysis: cluster analysis. a) Attribution of the PMF solutions to the clusters. b) Optimal PMF solutions. We initiated the cluster analysis 100 times; the selected solutions were those attributed to the optimal clusters (#1-#3) more than 95% of the cluster analyses. We note that although cluster 1 and 3 were chosen among the optimal clusters (clusters #1-#3), many of the PMF runs belonging to these clusters were attributed to suboptimal clusters (#4-#5) more than 5% of the time. Not surprisingly clusters 1 and 3 show better COA diurnal cycle correlations with the suboptimal clusters than cluster 2 (Table S3).

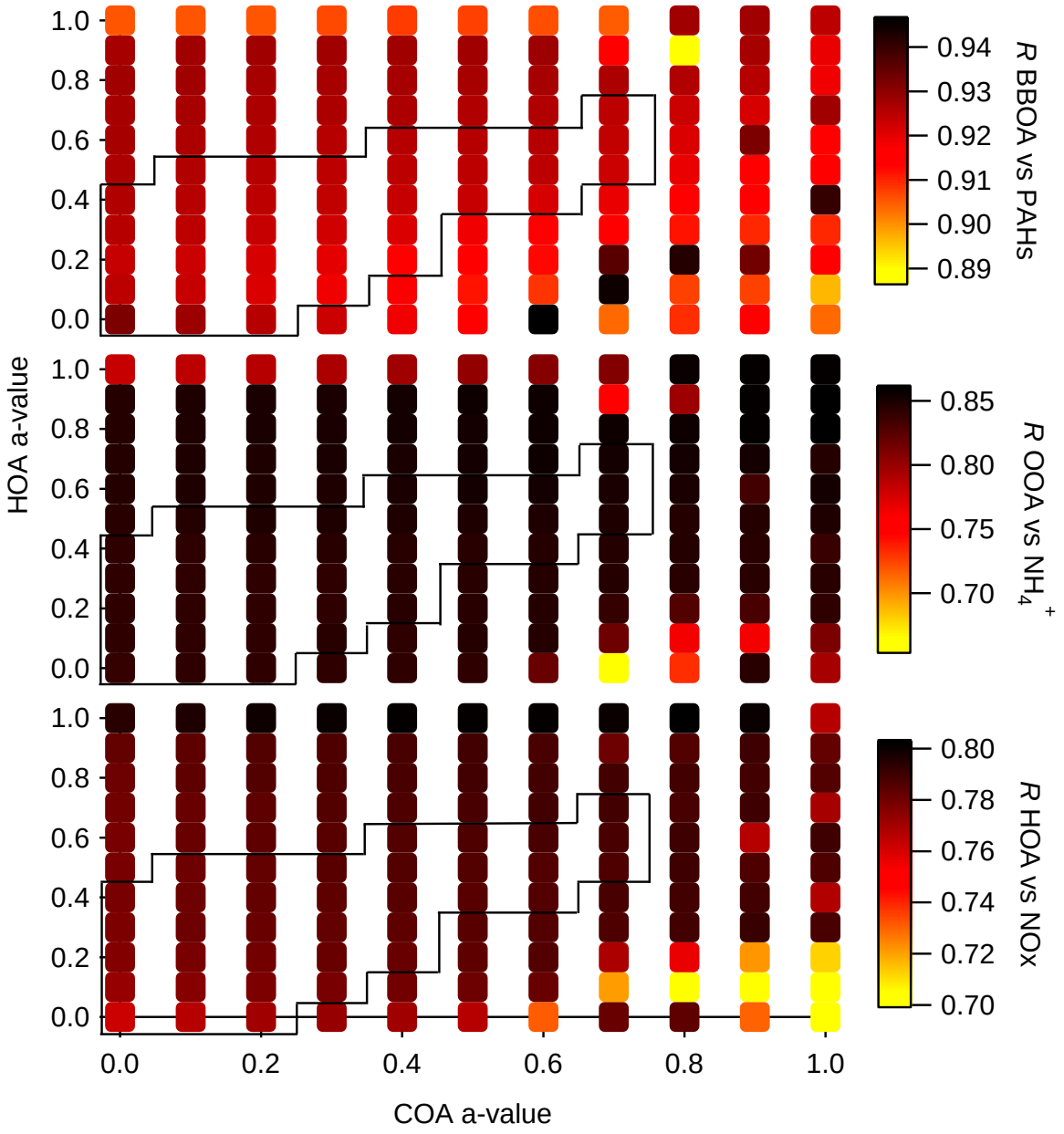


Figure S9. *a*-value sensitivity analysis diagnostics: correlations of factor profiles and corresponding tracers as a function of COA and HOA *a*-values.

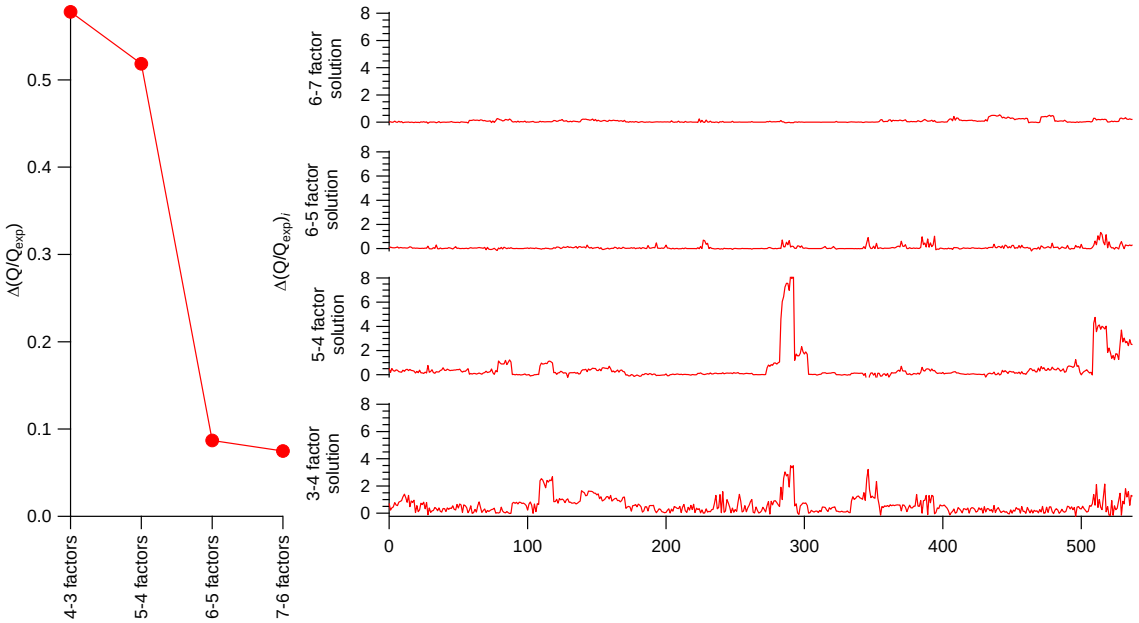


Figure S10. Offline-AMS PMF $\Delta(Q/Q_{exp})$ analysis. In this study, a large Q/Q_{exp} decrease could be observed up to 5 factors. Increasing the number of factors Q/Q_{exp} leads to smaller increase in the explained variability. The newly separated OOA factors could not be attributed to specific aerosol sources/processes.

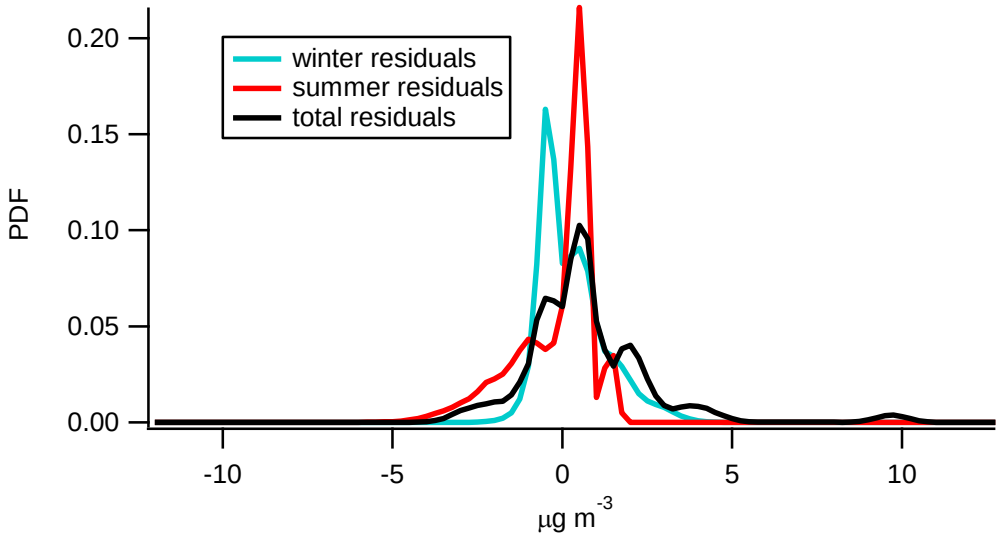


Figure S11. Probability density functions of the OC residuals from R_Z sensitivity analysis (Eq. 11).

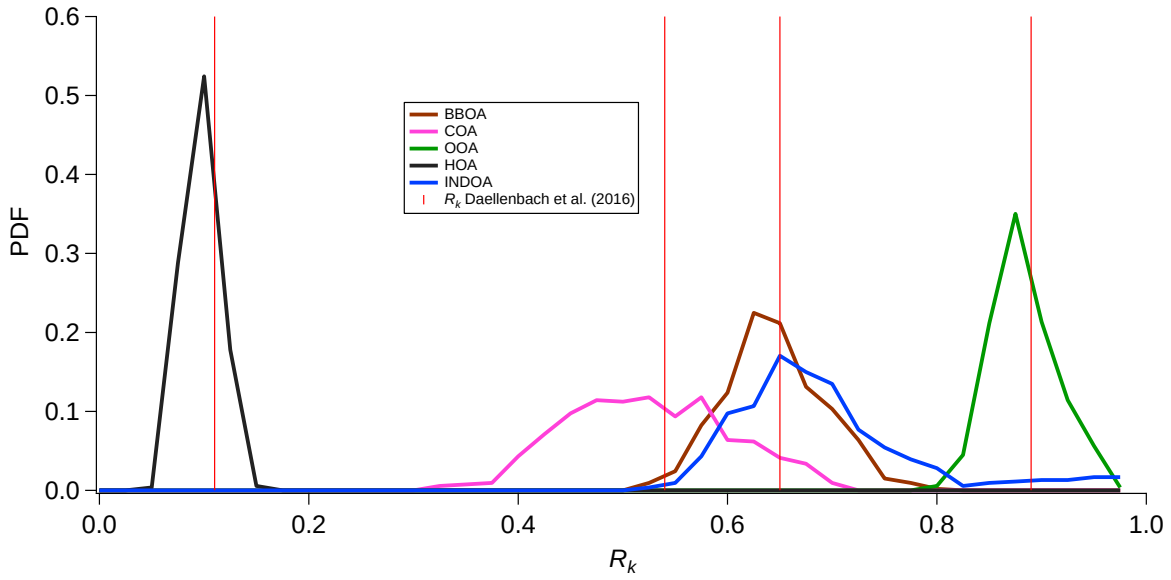


Figure S12. Factor recoveries: probability density functions. Vertical sticks represent the recoveries determined by Daellenbach et al. (2016). Estimated recoveries: $R_{HOA,med}$ = 0.11 (1st quartile 0.10, 3rd quartile 0.12); $R_{BBOA,med}$ = 0.65 (1st quartile 0.63, 3rd quartile 0.69); $R_{COA,med}$ = 0.53 (1st quartile 0.48, 3rd quartile 0.59); $R_{OOA,med}$ = 0.89 (1st quartile 0.87, 3rd quartile 0.91).

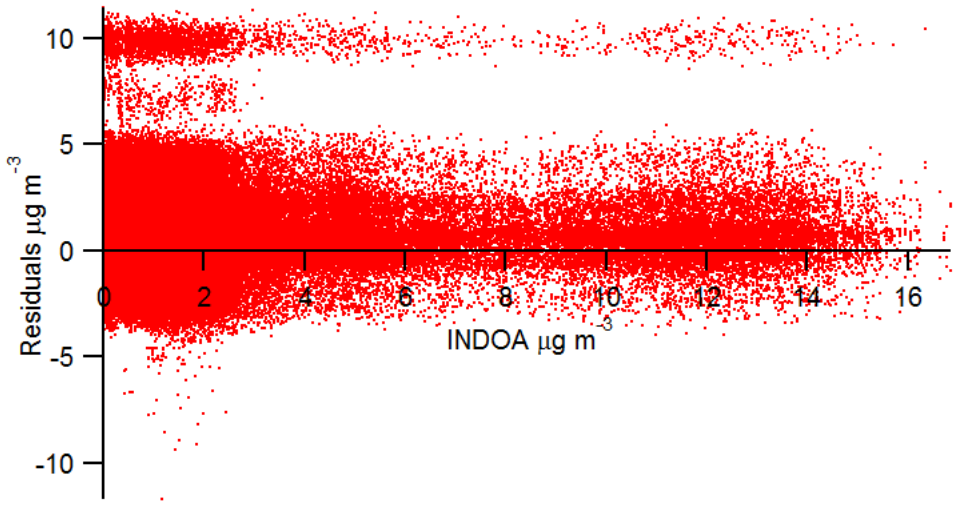


Figure S13. Offline-AMS: INDOA vs residuals concentrations (calculated according to Eq. 8) scatter plot. Residuals > 6 µg m^{-3} represented < 2% of the points and were associated to 25 November, where the C bulk extraction efficiency (Bulk EE = WSOC:OC) was estimated at 22%, representing an outlier in comparison to the median Bulk EE = 0.61 (1st quartile = 0.54; 3rd quartile = 0.71).

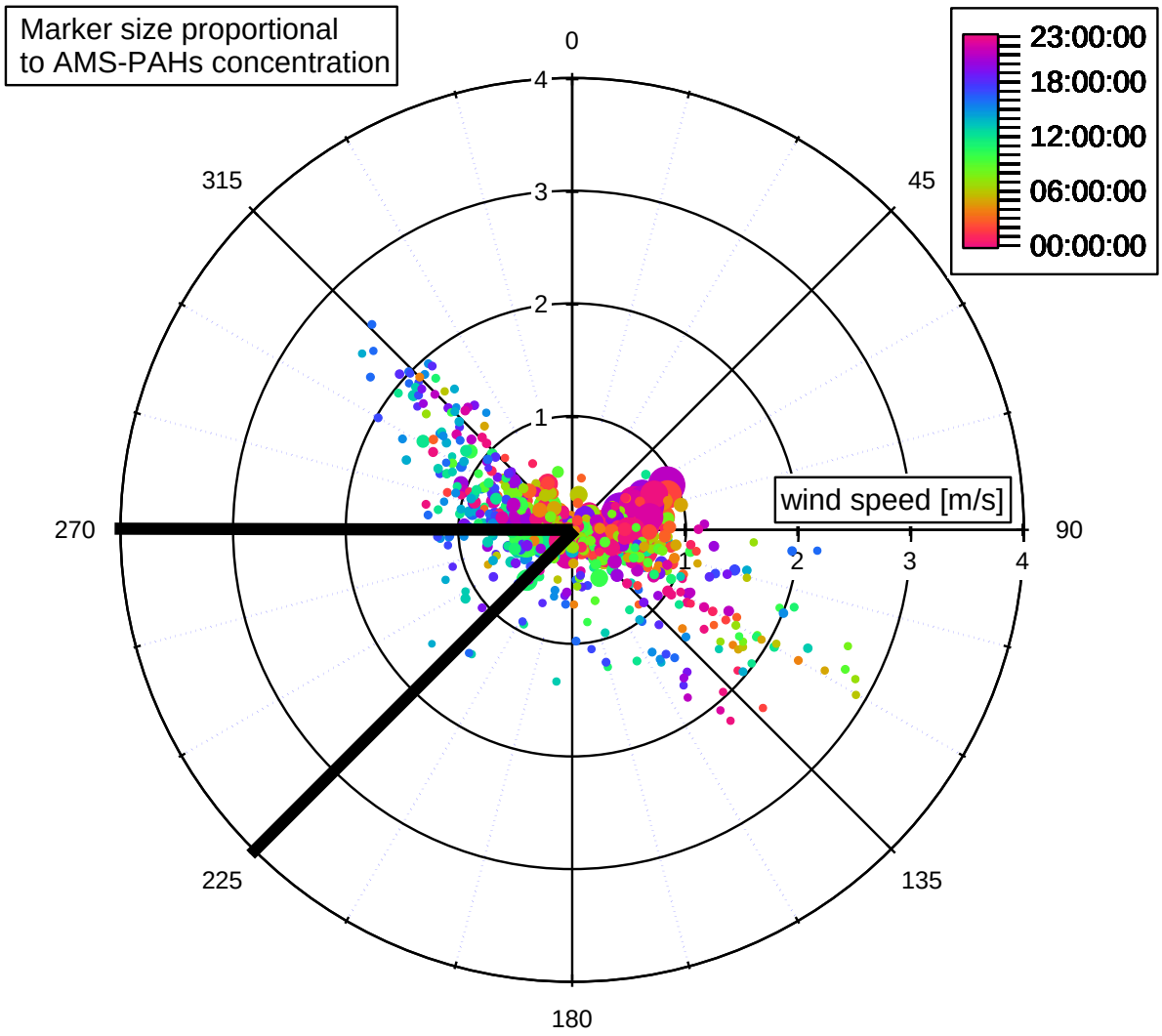


Figure S14. Online-AMS: AMS-PAHs rose plot. Color code denotes the hour of the day; marker size is proportional to the AMS-PAHs concentration. The distance from the center is proportional to the wind speed. El Haddad et al. (2013) revealed industrial-related emissions associated to wind direction from W/SW (225°-270°).

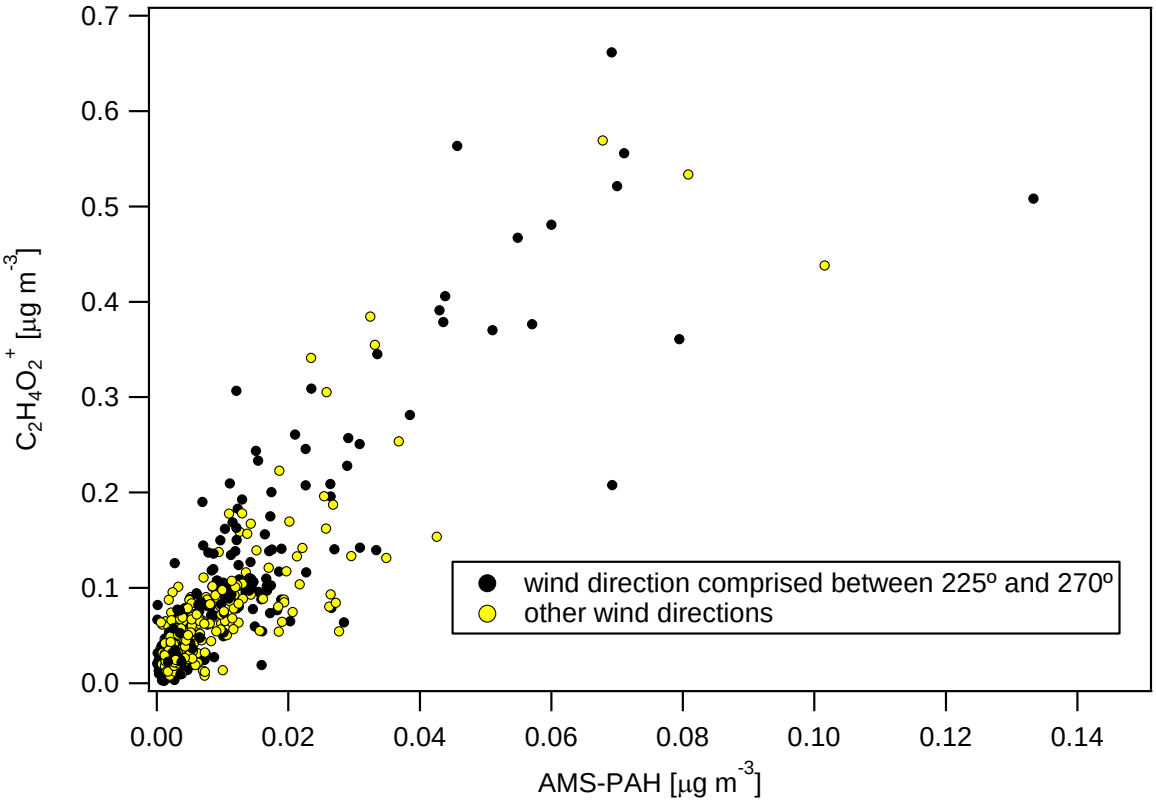
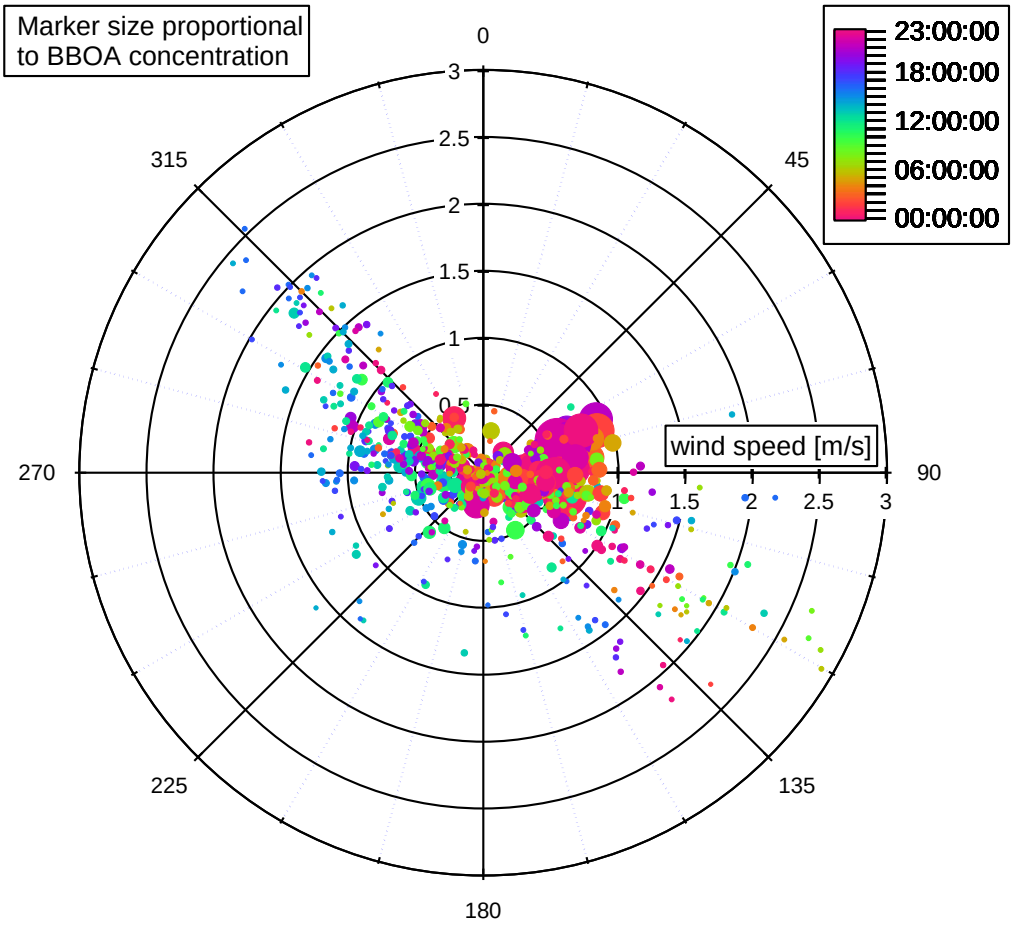
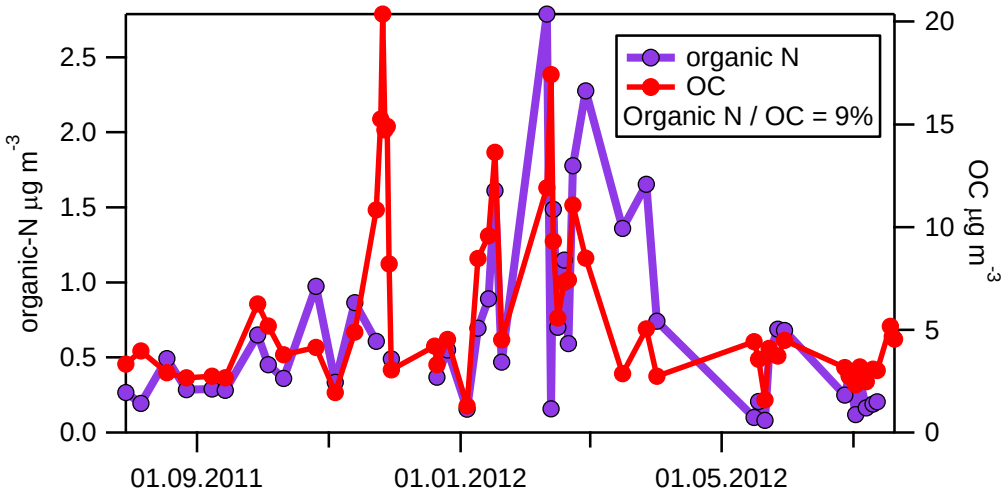


Figure S15. $C_2H_4O_2^+$ vs AMS-PAHs scatter plot.



1

2 Figure S16. Online-AMS: BBOA rose plot. Color code denotes the hour of the day, marker
3 size is proportional to the BBOA concentration. The distance from the center is proportional
4 to the wind speed.



5

6 Figure S17. Organic-N and OC time series.

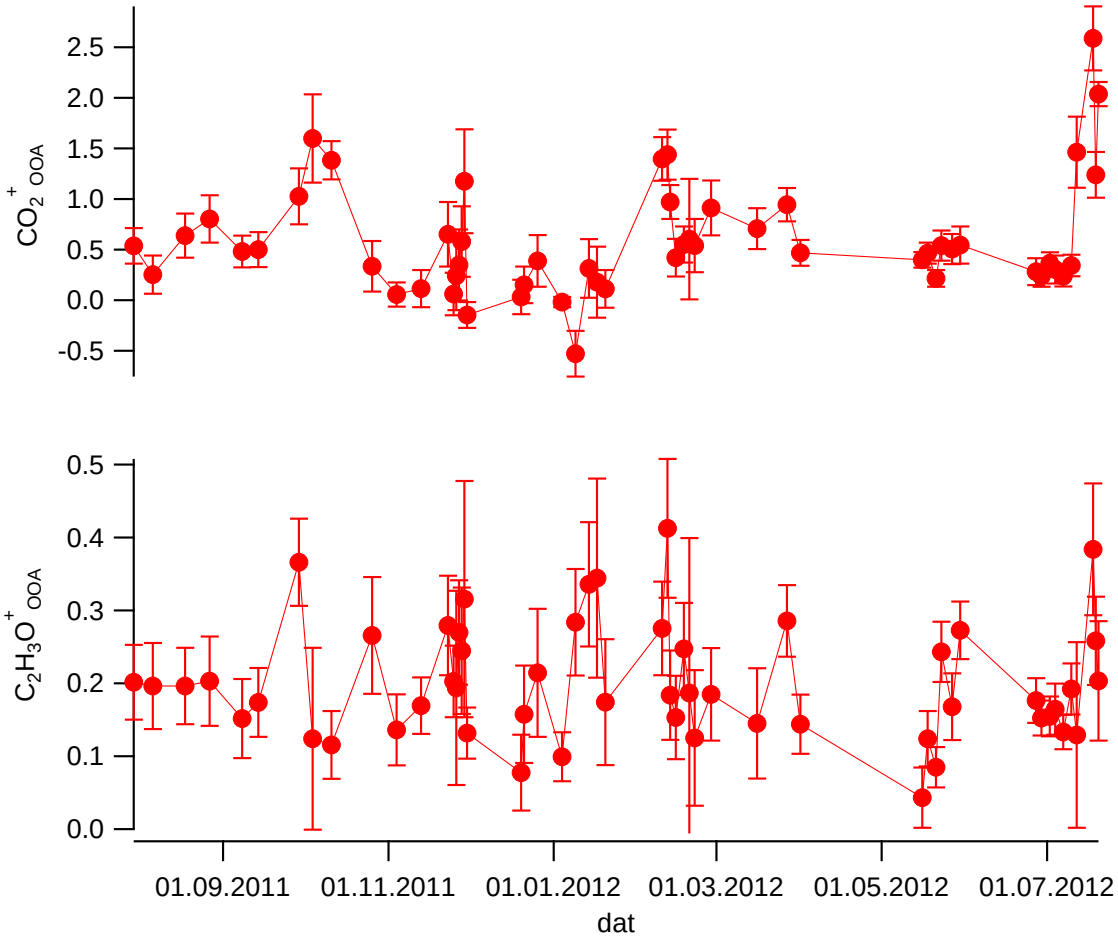


Figure S18. $C_2H_3O^+_{OOA}$ and $CO_2^+_{OOA}$ yearly cycle.

Comparison of online-AMS and offline-AMS mass spectra.

Offline-AMS source apportionment was conducted on water soluble OA AMS spectra. In this section we compare the online- and offline-AMS factor profiles obtained from PMF analysis. In general the measured water soluble AMS spectra show a higher O:C ratio (winter median = 0.51) than online-AMS spectra (median = 0.44). This is also indirectly confirmed by the high factor recoveries for the factors characterized by high O:C ratios (e.g. OOA and BBOA), while lower recoveries were estimated for the factors associated to low O:C ratios (e.g. COA and HOA, Fig. S13), indicating that offline-AMS accesses the most oxidized (i.e. water-soluble) OA fraction, consistent with Daellenbach et al., 2016. Table S4 reports the cosine similarities between the offline-AMS PMF factor profiles and the average reference online-AMS profiles (Crippa et al., 2013b; Mohr et al., 2012; 2009, Bruns et al., 2015; Docherty et al., 2011; Setyan et al., 2012; He et al., 2010).

Table S4. Cosine similarity between offline-AMS PMF factors and average online-AMS reference profiles (Crippa et al., 2013b; Mohr et al., 2012; 2009; Bruns et al., 2015; Docherty et al., 2011; Setyan et al., 2012; He et al., 2010). The threshold cosine similarity indicates the minimum cosine similarity value which defines a mass spectrum as not statistically different from the average online-AMS reference profiles. Values highlighted in red indicate offline-AMS factor profiles not statistically different from online-AMS reference spectra. Cosine similarities were calculated for all the retained offline-AMS PMF spectra. The values reported in the table represent the average cosine similarity; the corresponding uncertainty is reported in parenthesis.

Threshold cosine similarity			Offline-AMS factor profiles		
			HOA	COA	BBOA
0.957	Online-AMS average reference spectra	HOA	0.87 (1)	0.87 (1)	0.32 (1)
0.91		COA	0.931 (3)	0.933 (9)	0.46 (2)
0.89		BBOA	0.479 (7)	0.48 (1)	0.87 (2)

As expected from the relatively high recovery (median 0.65), the water soluble fraction of BBOA shows a mass spectral fingerprint not statistically different from the BBOA online-AMS reference spectra, although more oxidized (O:C = 0.54 for offline-AMS, and 0.35 for online-AMS). Despite the different degree of oxidation, the two spectra were not considered as different within our uncertainty due to the large variability of the BBOA AMS spectra reported in literature. As already mentioned, this variability mostly derives from the different burning conditions and fuels. In the same way also COA offline-AMS fingerprint is not statistically different from the online-AMS COA spectra. By contrast HOA, that has the lowest recoveries among the separated offline-AMS factors, shows a fingerprint which is statistically different than the HOA online-AMS reference spectrum, moreover the offline-AMS HOA fingerprint is statistically not different from the COA online-AMS reference spectrum. This is due to the higher HOA *a*-values associated with the accepted solutions (0.5, 0.9, 1) in comparison to the COA *a*-values (0.2, 0.3, 0.4, 0.8), which therefore enabled a significant variation of the mass spectrum . The resemblance of HOA and COA water-soluble spectra is due to the lower water solubility of saturated hydrocarbons in comparison to unsaturated hydrocarbons (Daellenbach et al. 2016). While online-AMS HOA literature spectra are characterized by similar values of saturated and unsaturated hydrocarbon fragments, the online-AMS COA reference spectra show higher values for saturated

hydrocarbon fragments in comparison to unsaturated ones (Mohr et al., 2009). The lower water solubility of saturated hydrocarbons therefore leads to similar COA and HOA fragmentation fingerprints for the water soluble spectra.